AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions of claims in the application:

Listing of Claims:

1-68. (Canceled)

69. (Currently Amended) A computer-implemented system for automatically categorizing unknown incoming data and a category visualization (CV) system that displays a graphic representation of each category as a hierarchical map, comprising:

a node corresponding to each base category;

nodes corresponding to combinations of similar categories;

a leaf node corresponding to a base category, the leaf node is positioned as a cluster of nodes at a lowest level of the hierarchy wherein combinations of similar categories are positioned on top of the leaf node, forming successively higher levels of the hierarchy;

a root node corresponding to a category that contains all records in a collection, the root node forms top of the hierarchy;

a non-leaf node corresponding to each combined category, wherein similar base categories are combined into a combined category; ~~and~~

wherein each non-leaf node has two arcs that connect the non-leaf node to two nodes corresponding to sub-categories of the combined category; and

wherein if a node is selected, the system displays additional information about corresponding category, the additional information is at least one of number of records in the category or characteristic attributes of the category.

70. (Previously Presented) The system of claim 69, wherein the base category is a category identified by a categorization process (classification and clustering).

71. (Previously Presented) The system of claim 69, wherein the combined category is assigned the records of two or more base categories.

72. (Cancelled)

73. (Currently Amended) The system of claim 69 [[72]], wherein the additional information further comprises characteristic and discriminating information such as attribute-value discrimination, attribute-value discrimination refers to how well the value of an attribute distinguishes the records of one category from the records of another category.

74. (Currently Amended) The system of claim 73, ~~wherein~~ the attribute-value discrimination is determined by employing the following equation:

$$discrim(x_i|G_1,G_2)=(p(x_i|G_1)-p(x_i|G_2))\log\frac{p(x_i|G_1)}{p(x_i|G_2)}+(p(x_i|G_2)-p(x_i|G_1))\log\frac{1-p(x_i|G_1)}{1-p(x_i|G_2)}$$

where $discrim(x_i|G_1,G_2)$ is the measurement of how well the value of an attribute distinguishes the records of one combined category from the records of another combined category,

G_1 is the first combined category,

G_2 is the second combined category,

x_i is the records in one of the combined categories,

$p(x_i|G_1)$ is the probability that a record containing specific attributes is in combined category G_1, and

$p(x_i|G_2)$ is the probability that a record containing specific attributes is in combined category G_2.

75. (Previously Presented) The system of claim 69, wherein if an arc is selected, the system displays information relating to categories connected by the arc, such as similarity value for the connected categories.

76. (Previously Presented) The system of claim 75, wherein similarity value refers to a rating of the differences between attribute values of records in one category and attribute values of records in another category, a high value for similarity indicates that there is little difference between the records in the two categories.

77. (Currently Amended) The system of claim 76, wherein the similarity value for a pair of base categories is determined by employing the following equation:

$$dist(h_1, h_2) = \sum_{x_1,\ldots,x_m} (p(x_1,\ldots,x_m|h_1) - p(x_1,\ldots,x_m|h_2)) \log \frac{p(x_1,\ldots,x_m|h_1)}{p(x_1,\ldots,x_m|h_2)}$$

where $dist(h_1, h_2)$ is the distance and similarity between two categories,

$x_1, \ldots, x_m$ is the attribute values,

h_1, h_2 is a count of a total number of records in categories 1 and 2,

$p(x_1, \ldots, x_m|h_1)$ is a conditional probability that a record has attribute values $x_1, \ldots x_m$ given that it is a record from category 1, and

$p(x_1, \ldots, x_m|h_2)$ is a conditional probability that a record has attribute values $x_1, \ldots x_m$ given that it is a record from category 2.

78. (Currently Amended) The system of claim 76, wherein the similarity for a pair of base categories is determined by employing the following equation:

$$dist(h_1, h_2) = \sum_{1} \sum_{x_i} (p(x_i|h_1) - p(x_i|h_2)) \log \frac{p(x_i|h_1)}{p(x_i|h_2)}$$

where $dist(h_1, h_2)$ is the distance and similarity between two categories,

x_i is the attribute values,

h_1, h_2 is a count of a total number of records in categories 1 and 2,

$p(x_i|h_1)$ is a conditional probability that a record has attribute values x_i given that it is a record from category 1, and

$p(x_i|h_2)$ is a conditional probability that a record has attribute values x_i given that it is a record from category 2.

79. (Currently Amended) The system of claim 76, wherein the similarity for two combined categories is determined by employing the following equation:

$$dist(G_1, G_2) = \sum_{h_j \in G_1, h_k \in G_2} (p(h_j)p(h_k)dist(h_j)p(h_j, h_k)$$

where $dist(G_1, G_2)$ is the distance and similarity between two combined categories,

G_1 is the first combined category,

G_2 is the second combined category,

h_j, h_k is a count of a total number of records in combined categories 1 and 2, and

$p(h_j)p(h_k)$ is a probability that a record is in each of the combined categories.

80. (Currently Amended) The system of claim 76, wherein the similarity for two combined categories is determined by employing the following equation:

$$dist(G_1, G_2) = \min\{dist(h_j)(h_k) \mid h_j \in G_1, h_k \in G_2\}$$

where $dist(G_1, G_2)$ is the minimum distance between two combined categories,

G_1 is the first combined category,

G_2 is the second combined category, and

h_j, h_k is a count of a total number of records in combined categories 1 and 2.

81. (Currently Amended) The system of claim 76, wherein the similarity for two combined categories is determined by employing the following equation:

$$dist(G_1, G_2) = \max\{dist(h_j)(h_k) \mid h_j \in G_1, h_k \in G_2\}$$

where $dist(G_1, G_2)$ is the maximum distance between two combined categories,

G_1 is the first combined category,

G_2 is the second combined category, and

h_j, h_k is a count of a total number of records in combined categories 1 and 2.

82. (Previously Presented) The system of claim 69, wherein the graphic representation of each category is displayed as a decision tree, further comprising:

nodes that correspond to each attribute of the corresponding base categories; and

arcs that correspond to values of that attribute;

wherein each node, except the root node, represents a setting of attribute values as indicated by arcs in a path from a first node to the root node.

83. (Previously Presented) The system of claim 82, wherein the selection of a node, results in display of a probability for each category that a record in the category will have attribute settings that are represented by the path.

84. (Withdrawn) A CV system that displays a graphic representation of each category as a similarity graph, comprising:

a node corresponding to each category; and

an arc that connects similar nodes;

wherein a similarity threshold is selected and arcs are displayed between nodes corresponding to pairs of nodes that are above the similarity threshold; and

wherein arcs between nodes are removed and added based upon changes to the similarity threshold.

85. (Withdrawn) The system of claim 84, wherein similar categories are combined.

86. (Withdrawn) The system of claim 84, wherein a category is split into sub-categories.

87. (Currently Amended) A computer-readable storage medium containing a plurality of categorized data records and a computer-implemented method of calculating and displaying a graphic representation of various characteristics and discriminating information for each category, comprising:

providing nodes that represent each base category;

providing nodes that represent combined categories, wherein combinations of similar categories are grouped together to form the combined categories;

utilizing a leaf node to form the bottom of the graphic representation;

utilizing a root node to form the top of the graphic representation;

connecting nodes representing sub-categories of a combined category via arcs;

combining the two base categories that are the most similar into a combined category; ~~and~~

repeating process of combining similar categories until one combined category represents all records in a collection; and

allowing a node to be selected, wherein the system displays additional information about corresponding category, the additional information is at least one of number of records in the category or characteristic attributes of the category.

88. (Previously Presented) The system of claim 87, further comprising de-emphasizing specific nodes and focusing on specific non-de-emphasized nodes.